A FUZZY LOGIC BASED EXPERT SYSTEM FOR QUALITY ASSURANCE OF DOCUMENT IMAGE COLLECTIONS

Roman Graf
AIT - Austrian Institute of Technology GmbH, Austria

Alexander Schindler
AIT - Austrian Institute of Technology GmbH, Austria

Reinhold Huber-Mörk
AIT - Austrian Institute of Technology GmbH, Austria

Abstract: Huge document image collections in digital libraries are prone to reduced quality and require automatic quality assurance. This paper presents an approach for bringing together information automatically aggregated from a quality assurance tool and expert knowledge related to digital preservation. The main contribution of this work is the definition of fuzzy expert rules and the application of fuzzy logic in order to support digital preservation experts in decision making. Page duplicate detection in document image collections is demonstrated in detail. Another contribution is a multi level analysis approach that comprises not only image processing, but also collection metadata aggregation e.g. file names, file size, creation date, possible inconsistency detection. This expert system supports planning for long term preservation and ensures quality of the digitized content. Our goal is to create a reliable inference engine and human maintainable conclusions from the output of an image processing tool that detects duplicates based on methods of computer vision. Another goal is to give a system at hand that supports digital document handling for teaching and education. We employ artificial intelligence technologies (i.e. fuzzy logic, expert rules) to emulate reasoning about the knowledge base similar to a human expert. A statistical analysis of the automatically extracted information from the image comparison tool and the qualitative analysis of the aggregated knowledge are presented in the evaluation part of the paper.

Keywords: digital preservation, fuzzy logic, quality assurance, image processing.

Introduction

By carrying out large-scale digitization projects digital libraries and archives are facing the challenge of maintaining sufficient quality in either the acquisition or the preservation of document image collections. Digital document image collections comprise millions of digitized artifacts (e.g. books, newspapers, journals) – each collection containing up to hundreds or thousands documents.

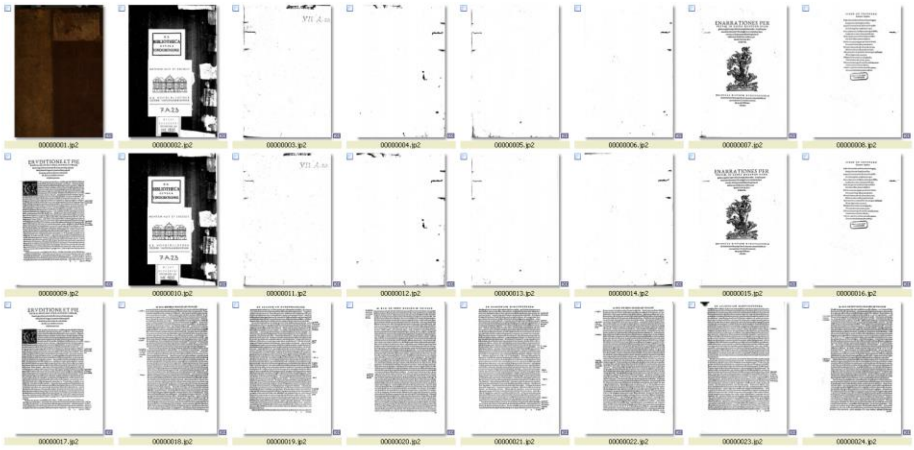

Figure 1. Sample of book scan sequence with a run of eight duplicated pages: images 10 to 17 are duplicates of images 2 to 9.

Manual maintenance and quality assurance of such collections is very time consuming resulting in high personel and storage costs. Therefore, automated solutions are required in order to manage and ensure quality of these collections. A typical quality assurance task of a library is to curate digitized books collections. For example, a digitization project running at the Austrian National Library produces digital images from books through an automatic scanning process without involvement of human interaction (see Figure 1). The produced digital collections are stored in the long term digital documents repository and are constantly merged with new versions applying image enhancement and OCR tools. At that point is important to select between the old and the new version of the associated documents due to the high cost of storage space. An automated approach based on an expert system should be able to make a decision about whether the documents should be replaced or removed. The data currently stored in digital collections is not structured, which additionally complicates the inspection process. Currently, collection holders do not have an automatic method able to detect duplicates and to remove them. A decision support system is required since human experts are not able to search and compare particular images among hundred thousand documents due to lack of time and concentration.

Our duplicate detection method for digital collection analysis is based on the matchbox tool [1, 2] that implements image comparison for digitized text documents. In this paper matchbox provides one of the fuzzy input variables of the fuzzy inference system. The main contribution of this paper is an evaluation of the matchbox tool in combination with a fuzzy expert system for the analysis of digital document collections. The output of an expert system is used for reasoning about analyzed data and for assessment w.r.t. duplicate detection and preservation risks. For the assessment of evaluation results we use ground data truth, manually created by experts from the Austrian National Library.

The paper is structured as follows: Section 2 gives an overview of related work and concepts. Section 3 explains the duplicate detection methods workflows and also covers image processing issues. Section 4 describes fuzzy modelling. Section 5 presents the experimental setup, applied methods and results interpretation. Section 6 concludes the paper and gives outlook on planned future work.

Related Work

Image processing techniques could be employed for quality assurance of digital content by replacing of a human expert regarding the decision-making process in a particular domain. The matchbox approach of duplicate detection employs structural similarity computation. This tool is based on local image descriptors, namely SIFT [3], in combination with a bag of visual words (BoW) dictionary representation. In general, several image retrieval and comparison approaches in large image collections make use of local image descriptors and visual dictionary methods to match or index visual information. Near duplicate detection of key frames using one-to-one matching of local descriptors was described for video data [4]. A BoW [5] derived from local descriptors was described as an efficient approach to near-duplicate video key frame retrieval [6]. Local descriptors were employed for the detection of near-duplicates in [7]. Note, that the use of optical character recognition, which is an obvious approach for the extraction of relevant information from text documents, is quite limited with respect to accuracy and flexibility [8]. The digital preservation employs computer vision techniques for different scenarios. Strodl et al [9] present the Planets [10] preservation planning methodology by an empirical evaluation of image scenarios. They demonstrate specific cases of recommendations for image content in four major National Libraries in Europe.

Fuzzy logic is one method that can be employed for modelling and reasoning based on an expert system. Prominent work in this field includes the rule-based system presented by Bernard [11]. It was designed for processing and power control in a power plant. A survey of the fuzzy logic controller (FLC) [13] evaluates the linguistic control methodologies, the derivation of the fuzzy control rules and an analysis of fuzzy reasoning mechanisms.In order to evaluate a control action the relevant parameters are measured. Actions are specified in rules. The fuzzy logic and inference engine is used to search through the knowledge base in order to identify those rules that are applicable. This approach is very similar to our Expert System organization, with the difference that we have a different modelling by membership functions, fuzzy standards and by fuzzy result that optionally can be defuzzified. An application of natural language words instead of numbers for computing and reasoning using fuzzy logic is described in [12]. The proposed Expert System makes use of this technique. This research is important since the FLC is used in our approach as a standard for fuzzy rules definition. The qualitative safety modelling in [14] is performed employing fuzzy IF - THEN rules. Compared to existing systems the proposed system is more efficient due to the use of more complex fuzzy rules. Simple IF - THEN rule engines are not well suited for dealing with aggregated quality analysis data having a level of uncertainty that is dependent on particular collection. A fuzzy-logic based approach may be more appropriately for the duplicate detection analysis. The provided Expert System deals directly with the linguistic terms commonly used in the digital preservation community for quality assessment of digital collection. Our research focuses on the development and representation of linguistic variables and subsequent calculation of their membership functions. These variables are then quantified using input numbers and fuzzy logic with the goal to decide, whether given document is a duplicate or not.

Duplicate Detection Process

Decision making process for quality assurance in digital preservation plays an important role and requires expertise in file formats and regular library processes. The search for such knowledge requires expertise and is time consuming. A consistent collection should not contain duplicates or ambiguous entries.

The matchbox tool uses the SIFT interest point detection and local feature descriptors [3], which have proven highly invariant to geometrical and radiometrical distortions [15] and were successful applied to a variety of problems in computer vision. In our approach we make use of a BoW of 1000 visual words created using k-means clustering of the SIFT descriptors extracted from all images in a given collection. This can become computationally very demanding. As a single scanned book page already contains a large number of descriptors, we applied preclustering of descriptors to each image. In contrast to a similar procedure [16], where all descriptors for all images of the same category are clustered independently using k-means method and subsequently appended to the BoW, we construct a list of clustered descriptors for each book page and cluster this list in a second step in order to obtain a dictionary for the whole book. The similarity score between two documents is obtained from the comparison of corresponding keyword frequency histograms followed by structural similarity comparison.

In order to identify duplicates we aggregate collection specific knowledge and analyze collections using the matchbox tool for extracting the metadata of a collection. Figure 2 demonstrates the duplicate detection workflow using the matchbox quality assurance workflow. A user may trigger either a complete collection analysis in one turn or performs collection analysis in separate steps. In the latter case Python scripts with different parameters can explicitly extract document features, create a visual dictionary according to the BoW method, create visual histograms based on the visual dictionary for each document and, finally, perform a pair-wise comparison of all documents. Subsequently a human expert should validate the provided list of duplicate candidates. The "all" path in the workflow means that all workflow steps will be executed per default. Additionally, duplicate candidates contained in a shortlist can be validated by structural similarity comparison based on SSIM [17], which requires additional computation time.

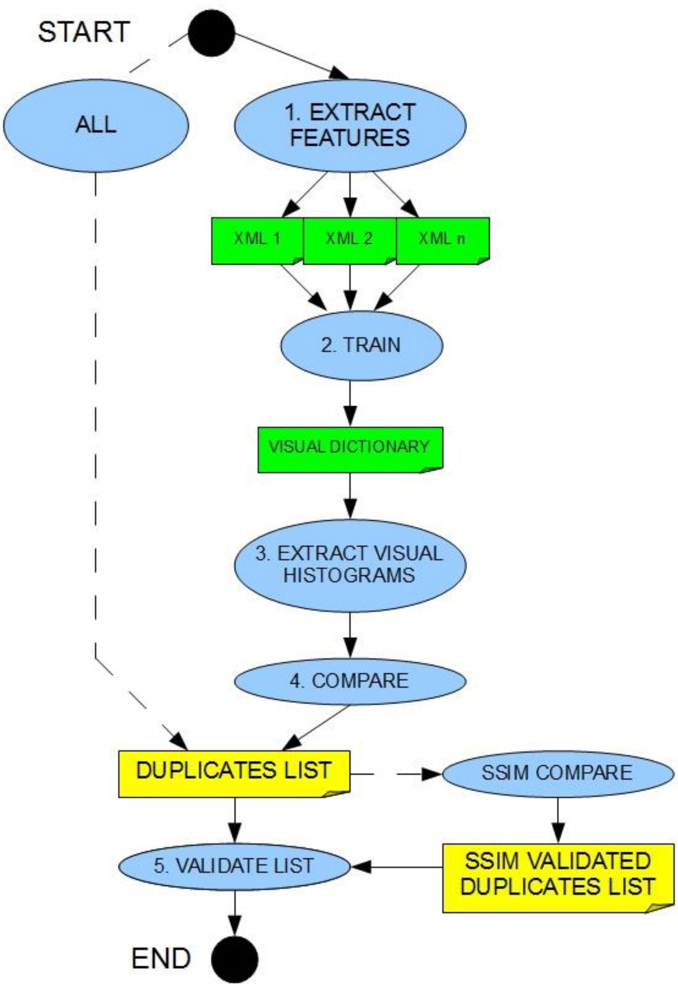

Figure 2. A workow for matchbox duplicate detection approach.

Fuzzy Modelling

In a fuzzy logic variables can take not only two states (true and false) but can have a not fixed thruth value in range from completly false (0) to completely true (1). Fuzzy logic also makes use of linguistic variables, in our case defined by digital preservation experts that can be presented by member functions of different complexity. The fuzzy variables describe expert rules and can be defined according to standard Fuzzy Control Logic (FCL). The advantage of this concept is that every expert can adjust linguistic variables, its member functions and thresholds according to preferences, requirements and policies of particular institutions.
In order to evaluate duplicates we determined a control system based on fuzzy logic that comprises eight inputs and one output presented in Figure 3.

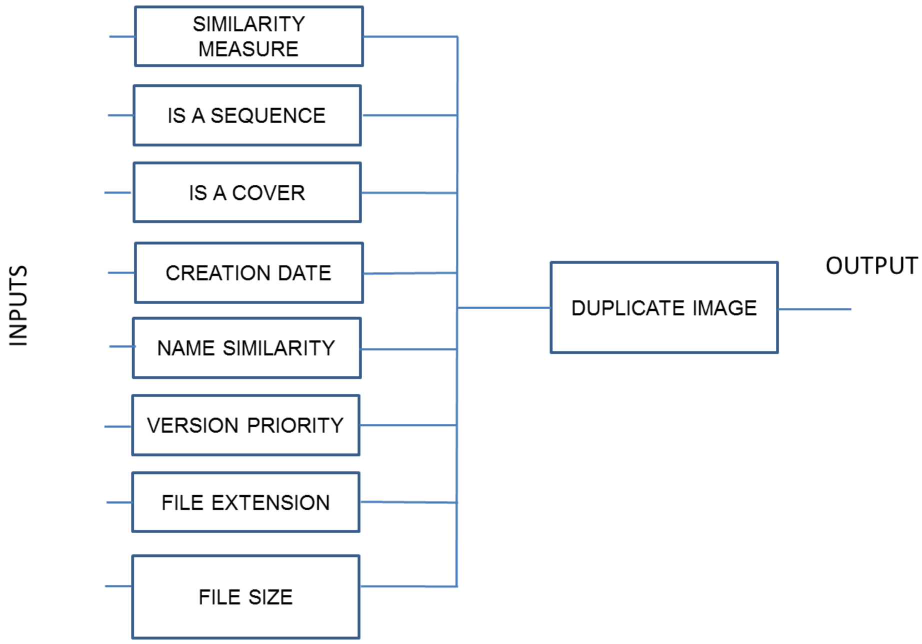

Figure 3. A control system based on fuzzy logic.

Input variables describe an input from the matchbox tool in "Similarity_Measure" variable and metadata from Phyton script presented by "Is_a_Sequence", "Is_a_Cover", "Creation_Date", "Name_Similarity", "Version_Priority", "File_Extension" and "File_Size" variables. The output is expected in a "Duplicate_Image" variable.
Fuzzy inference starts with fuzzification where numerical values are mapped to the membership functions associated with fuzzy variables. Then we apply logic defined in FCL to evaluate fuzzy rules and aggregate their output. These outputs are mapped to the output variables. Finally, these variables are mapped back to the numerical values. The example of duplicate evaluation is presented in a Figure 14 in evaluation part of this paper.

An input variable “Similarity_Measure” contains three membership functions flagged by the linguistic variables "No, Low and High", which are depicted in Figure 4. A corresponding graphical representation is shown in Figure 10. The values for these linguistic variables range from 0 to 1 and are coming from the matchbox tool. For simplicity we transform these values to percents. Therefore, structural similarity can be defined as high if its value matches in a range between 67 and 100 percent. In contrast low similarity values are between 14 and 88 percent. Finally values between 0 and 20 percent indicate that there is no structural similarity for analyzed image document.

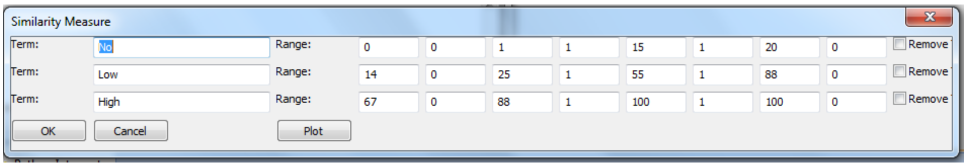

Figure 4. Definitions of fuzzy membership functions with associated values for input variable “Similarity_Measure”.

Three linquistic variables presented in Figures 5 and 11 were defined for “Is_a_Sequence” variable. The reason for that is the complexity of the sequence definition in a digital collection. For different collections sequence size can vary depending on digitization hardware, software and quality. The sequence size here means how many pages should be rescanned after an error in a scaning process. The quality here means the image resolution and the ability of scanning system for automated error detection.

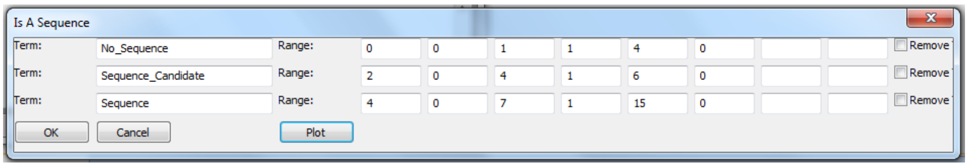

Figure 5. Definitions of fuzzy membership functions with associated values for input variable “Is_a_Sequence”.

The “Name_Similarity” and “File_Size” variables depicted in Figures 6, 7, and 12 are similar in their complexity and have two simple member functions that will be converted to Boolean values in the evaluation part.

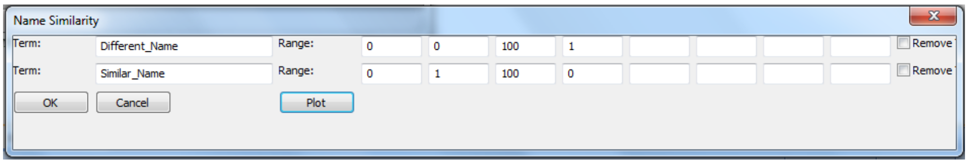

Figure 6. Definitions of fuzzy membership functions with associated values for input variable “Name_Similarity”.

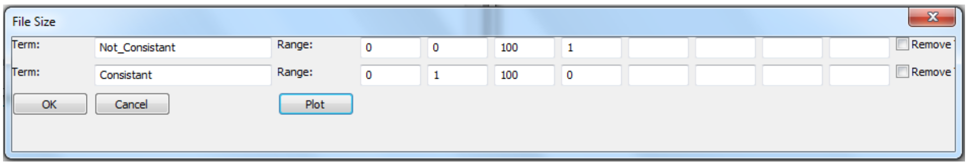

Figure 7. Definitions of fuzzy membership functions with associated values for input variable “File_Size”.

In the “Is_a_Cover” variable settings of member functions presented in Figures 8 and 13 were defined making use of an expert knowledge. It was observed, that the front and back matter of books differs from the main body w.r.t. the book cover pages, which are often textured in historical books, and series of empty or sparsely filled pages. The suggested rule avoids that those pages are falsely categorized as duplicated pages.

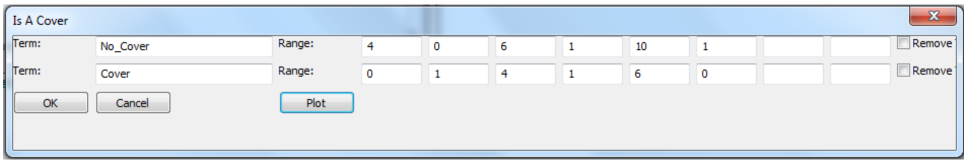

Figure 8. Definitions of fuzzy membership functions with associated values for input variable "Is_a_Cover".

The member functions of the "Version_Priority" and "File_Extension" variables depicted in Figure 9 have only two states: 0 or 1. These variables are also presented as a Boolean value in evaluation part.

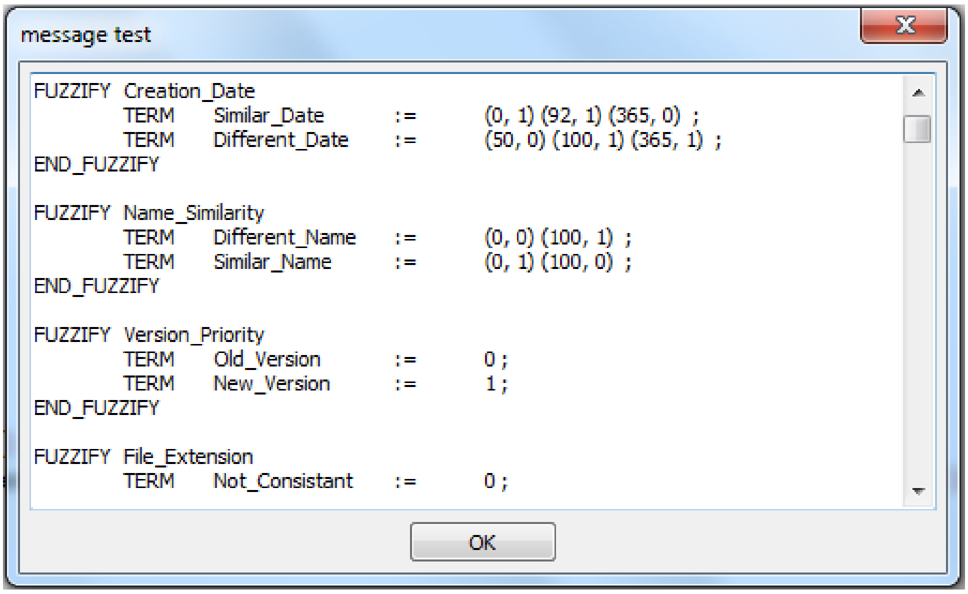

Figure 9. Excerpt from the FCL configuration file with definitions of fuzzy membership functions for input variables "Creation_Date", "Version_Priority" and "File_Extraction".

A fuzzy set for “Similarity_Measure” and its membership functions “No”, “Low” and “High” can be described by Equations 1-4.

$$(U, m) = \left\{ \frac{m(x_{NO})}{x_{NO}}, \frac{m(x_{LOW})}{x_{LOW}}, \frac{m(x_{HIGH})}{x_{HIGH}} \right\}, x \in U \qquad (1)$$

$$m(x_{NO}) = \begin{cases} 1, & if\ 0 < x \leq 15 \\ -\dfrac{x}{5} + 4, & if\ 15 < x \leq 20 \end{cases} \qquad (2)$$

$$m(x_{LOW}) = \begin{cases} \dfrac{x}{11} - 1{,}27, & if\ 14 < x \leq 25 \\ 1, & if\ 25 < x \leq 55 \\ -\dfrac{x}{33} + 1{,}66, & if\ 55 < x \leq 88 \end{cases} \qquad (3)$$

$$m(x_{HIGH}) = \begin{cases} \dfrac{x}{11} - 0{,}69, & if\ 67 < x \leq 88 \\ 1, & if\ 88 < x \leq 100 \end{cases} \qquad (4)$$

Where (U,m) denotes a fuzzy set whose elements x have a grade of membership - not included (m(x) = 0), fully included (m(x) = 1) or x is defined by membership function m(x). The Figures 10 to 13 depict graphical representation of previously defined fuzzy rules and its membership functions.

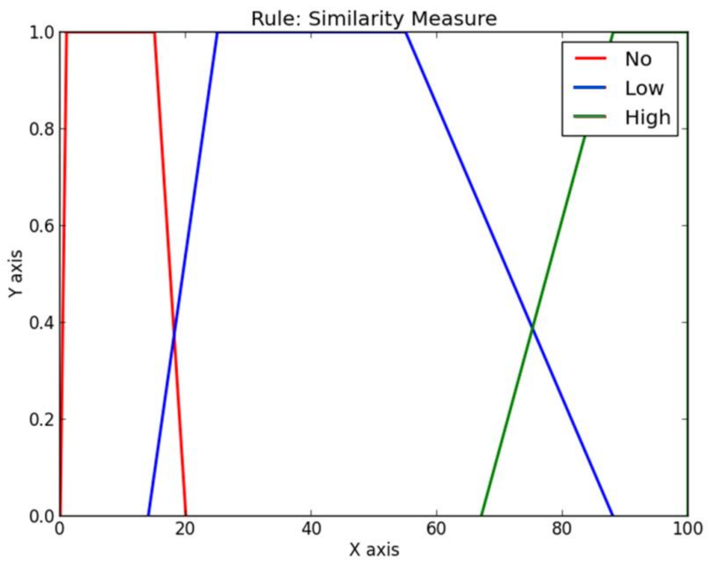

Figure 10. Plot of membership functions for similarity fuzzification.

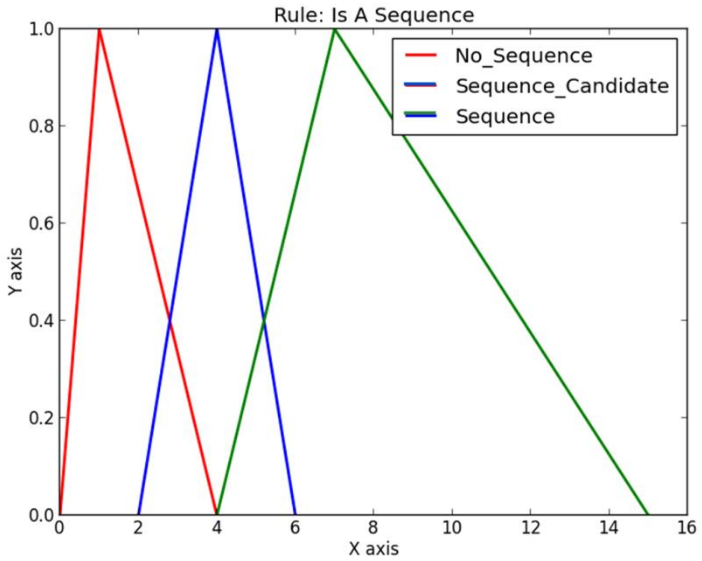

Figure 11. Plot of membership functions for sequence fuzzification.

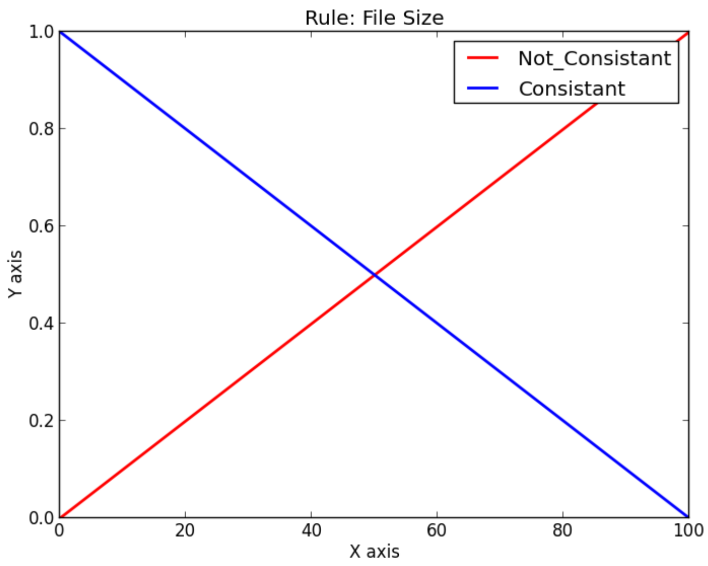

Figure 12. Plot of membership functions for file size fuzzification.

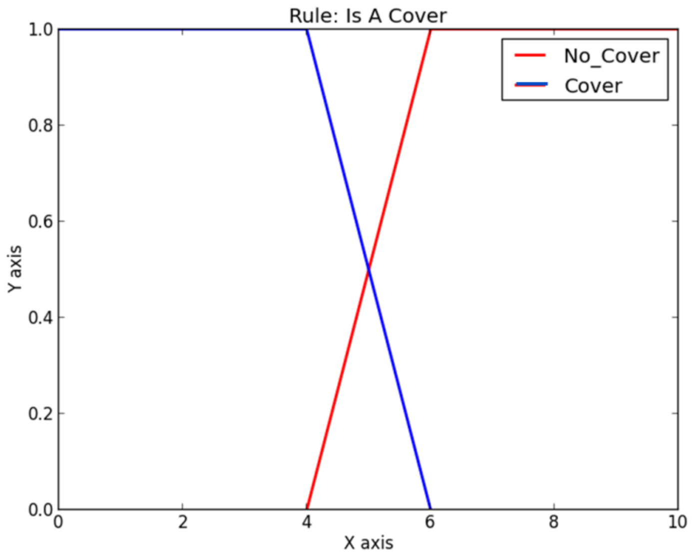

Figure 13. Plot of membership functions for cover fuzzification.

Experimental Results and its Interpretation

The goal of evaluation is a calculation of fuzzy rules that comprise image processing methods and metadata analysis for digital collection cleaning by removing of duplicates. Another goal is to adjust the accuracy of the expert rules, fuzzy input variables and its member functions for robust analysis. Additionally, a quantitative overview of evaluated data and methods characteristics is delivered. The considered collections with identifiers Z151671106, Z137274000, Z13721930X, Z137114501 and Z151698604 are provided by the Austrian National Library. Associated manually created ground truth was available. Provided that an automatic fuzzy logic based approach performs at least comparable to a human expert the suggested method would be a significant improvement over a manual analysis and could be used by experts in digital libraries and archives. Evaluation takes place on an Intel Core i7-3520M 2.66GHz computer using Java, Python and C++ languages on Windows OS. We evaluate duplicate candidate pairs, metadata characteristics of collections and fuzzy inference system output.

Table 1. Evaluation results for selected digital collections.

Barcode	Pages	#Sequences	Creation Date	File Name	File Size	Extension	#Pairs	Precision	Recall	Groundtruth	TP	FN	FP
Z151671106	482	2	T	T	F	T	16	0,9375	0,6818	22	15	7	1
Z137274000	292	1	T	T	F	T	19	1,0000	0,6316	19	12	7	0
Z13721930X	986	2	T	T	F	T	6	0,0784	1,0000	4	4	0	47
Z137114501	416	2	T	T	F	T	17	1,0000	0,8125	16	13	3	6
Z151698604	490	3	F	T	F	T	15	0,8947	0,7727	22	17	5	2

In the Table 1 are presented evaluation results for five selected digital collections. The “Pages”, “#Sequences”, “Creation Date”, “File Name”, “File Size” and “Extension” columns show results of the metadata analysis performed by the fuzzy inference engine described in previous chapters. These parameters describe collection documents by its overall characteristics. Whereas character “T” stands for true and character “F” stands for false. The columns “Precision”, “Recall”, “Ground truth”, “True Positives” (TP), “False Negatives” (FN) and “False Positives” (FP) present the subsequent quantitative analysis of the aggregated data. Though, the “File Size” calculation was not efficient for presented experiment (size varies between 3KB and 918KB for collection Z151671106) this condition could be applied for another type of collections where size estimation could make sense. Some detected duplicate pairs ([2,985] and [3,984] in Z13721930X collection and [3,414] in Z137114501 collection) due to inconsistent distances are not belonging to a sequence and therefore according to definition of the "Is_a_Sequence" rule were not marked as a duplicate. The creation date of documents in given collections is always the same for each collection. Only Z151698604 collection demonstrates small difference of 1. The file extensions are also the same. The file name has the same length (8 characters) and these are only numbers that continuously enumerate collection. The reason for that is that these collections are generated during the automated scanning process.

Table 2. Fuzzy inference engine results for collection Z13721930X.

Page Number	Duplicate Candidate	Distance	Is a Sequence	Is a Cover	Creation Date	File Name	File Size	Extension	Version Priority	Result
2	985	983	F	T	T	T	T	T	New	Is_Not_Duplicate
3	984	981	F	T	T	T	T	T	New	Is_Not_Duplicate
210	212	2	T	F	T	T	T	T	New	Is_ Duplicate
211	213	2	T	F	T	T	T	T	New	Is_ Duplicate
440	442	2	T	F	T	T	T	T	New	Is_ Duplicate
441	443	2	T	F	T	T	T	T	New	Is_ Duplicate

Table 2 depicts the fuzzy inference engine results for collection Z13721930X. The file sizes for given documents in this case are quite similar. For duplicate candidate 984 it is 455KB and 985 - 464KB, 2 – 418KB, 3 – 432KB, 210 – 285KB, 211 – 295KB, 212 – 288KB, 214 – 263KB, 440 – 271KB, 441 – 309KB, 442 – 304KB, 443 – 272KB, respectively. This observation demonstrates that despite the duplicate documents semantically are the same, they are still different image files. With small difference assumption of about 10% these values are similar according to each other. That small difference of duplicated documents should be taken into account in definition of associated fuzzy member functions. The pages 2, 3, 984 and 985 were correctly marked as a cover because of their position at the beginning and the end of the book. The structure of these documents is very similar but not identical.

Table 3. Fuzzy inference engine results for collection Z137114501.

Page Number	Duplicate Candidate	Distance	Is a Sequence	Is a Cover	Creation Date	File Name	File Size	Extension	Version Priority	Result
3	414	413	F	T	T	T	T	T	New	Is_Not_Duplicate
80	88	8	T	F	T	T	T	T	New	Is_Duplicate
81	89	8	T	F	T	T	T	T	New	Is_Duplicate
82	90	8	T	F	T	T	T	T	New	Is_Duplicate
83	91	8	T	F	T	T	T	T	New	Is_Duplicate
84	92	8	T	F	T	T	T	T	New	Is_Duplicate
85	93	8	T	F	T	T	T	T	New	Is_Duplicate
86	94	8	T	F	T	T	T	T	New	Is_Duplicate
87	95	8	T	F	T	T	T	T	New	Is_Duplicate
120	130	10	T	F	T	T	T	T	New	Is_Duplicate
121	131	10	T	F	T	T	T	T	New	Is_Duplicate
122	132	10	T	F	T	T	T	T	New	Is_Duplicate
123	133	10	T	F	T	T	T	T	New	Is_Duplicate
124	134	10	T	F	T	T	T	T	New	Is_Duplicate
125	135	10	T	F	T	T	T	T	New	Is_Duplicate
126	136	10	T	F	T	T	T	T	New	Is_Duplicate
127	137	10	T	F	T	T	T	T	New	Is_Duplicate

In the Table 3 are presented two sequences with different distances and one cover page. Figure 1 demonstrates calculated degree of truth that ranges between 0 and 1 that analyzed document 00000088.jp2 from collection Z137114501 is a duplicate.

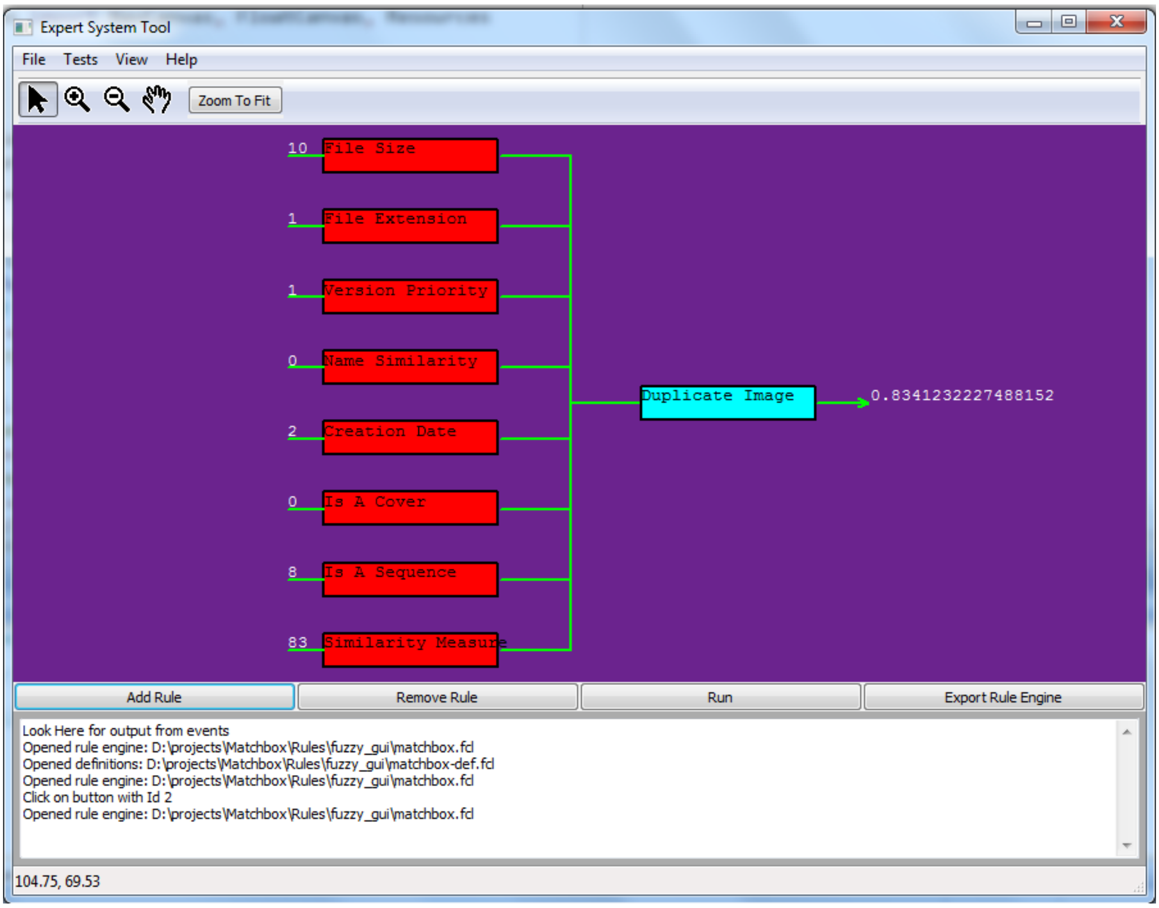

Figure 14. A fuzzy inference system for calculation of the truth degree for being a duplicate candidate for the document 00000088.jp2 in collection Z137114501.

The expert system starts duplicate identification with numerical input values on the left side where white numbers are shown. For "Similarity_Measure" the input value is 0.83. Fuzzifying this value we map it to the associated numerical value using the FCL input variables definition. For the input variable "Similarity_Measure" associated linguistic term is “High”. Aggregating all rule outputs we defuzzify the output value of the total confidence level for “Is_a_Duplicate” and obtain a value of 0.834123.
The document 00000088.jp2 is a duplicate with degree of truth 0.834. A structural similarity of this file with the document 00000080.jp2 of the same collection is 0.83. The document belongs to a sequence of 8 documents and is not a cover. The corresponding document has approximately the same file size, the same file extension and creation date. The newer version of the document is a priority version and should replace old version.

Conclusion

In this paper we presented an approach for bringing together information automatically aggregated from the matchbox quality assurance tool and expert knowledge from the filed of digital preservation. Large document image collections in digital libraries require automatic quality assurance. The main contribution of this work is the definition of fuzzy input and output variables and application of a fuzzy inference system to support digital preservation experts in decision making for page duplicate detection in document image collections. Another contribution is a multi level quality analysis approach that comprises not only image processing, but also collection metadata aggregation e.g. file names, file size, creation date, possible inconsistency detection. The suggested expert system supports planning for long term preservation and ensures quality of the digitized content. We created a reliable inference engine that makes use of the output of an image processing tool matchbox, which detects duplicate candidates based on methods of computer vision.
During evaluation we adjusted thresholds for member functions of fuzzy input variables in order to make expert system more robust for new evaluations. Therefore, we gave a system at hand that supports digital document handling forlibraries and archives. In this work we employed artificial intelligence technologies (i.e. fuzzy logic, expert rules) to emulate reasoning about the aggregated knowledge base similar to a human expert. The evaluation demonstrates that given approach enables integration of complex rules, provides automatical decision support and helps in solving practical digital preservation issues like duplicate detection. The calculated fuzzy results and their linguistic interpretation provided by developed Expert System are about the reduction of uncertainty by quality analysis. The proposed system is unique for the given domain.
As future work we plan to increase the amount of the fuzzy input variables, to extend an Expert System with additional fuzzy input variables, to improve the accuracy of its member functions and to increase quality of the outputs.

Acknowledgments

This work was partially supported by the SCAPE Project.The SCAPE project is co-funded by the European Union under FP7 ICT-2009.4.1 (Grant Agreement number 270137).

References

[1] R. Huber-Mörk and A. Schindler. Quality assurance for document image collections in digital preservation. In Proc. of the 14th Intl. Conf. on ACIVS (ACIVS 2012), volume 7517 of LNCS, pages 108-119, Brno, Czech Republic, September 4-7 2012. Springer.
[2] R. Huber-Mörk, A. Schindler, and S. Schlarb. Duplicate deterction for quality assurcance of document image collections. In In iPRES 2012 - Proceedings of the 9th International Conference on Preservation of Digital Objects, pages 136-143, Toronto, Canada, October 1-5 2012.
[3] D. G. Lowe. Distinctive image features from scale-invariant keypoints. Int. J. of Comput. Vision, 60(2):91-110, 2004.
[4] W.-L. Zhao, C.-W. Ngo, H.-K. Tan, and X. Wu. Near-duplicate keyframe identification with interest point matching and pattern learning. Trans. Multi., 9(5):1037-1048, Aug. 2007.
[5] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In In Workshop on Statistical Learning in Computer Vision, ECCV, pages 1-22, 2004.
[6] X. Wu, W.-L. Zhao, and C.-W. Ngo. Near-duplicate keyframe retrieval with visual keywords and semantic context. In Proceedings of the 6th ACM international conference on Image and video retrieval, CIVR '07, pages 162-169, New York, NY, USA, 2007. ACM.
[7] Y. Ke, R. Sukthankar, and L. Huston. An efficient parts-based near-duplicate and sub-image retrieval system. In Proceedings of the 12th annual ACM international conference on Multimedia, MULTIMEDIA '04, pages 869-876, New York, NY, USA, 2004. ACM.
[8] J. van Beusekom, D. Keysers, F. Shafait, and T. Breuel. Distance measures for layout-based document image retrieval. In 2nd ICDIAL, 2006. DIAL '06, pages 231-242, April 2006.
[9] S. Strodl, C. Becker, R. Neumayer, and A. Rauber. How to choose a digital preservation strategy: evaluating a preservation planning procedure. In: JCDL 2007: Proceedings of the 2007 conference on digital libraries, pages 29-38, New York, NY, USA, 2007. ACM.
[10] S. Schlarb, E. Michaelar, M. Kaiser, A. Lindley, B. Aitken, S. Ross, and A. Jackson. A case study on performing a complex file-format migration experiment using the planets testbed. IS&T Archiving Conference, 7:58-63, 2010.
[11] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In Workshop on Statistical Learning in Computer Vision, ECCV, pages 1-22, 2004.
[12] Lee, C.C.: Fuzzy logic in control systems: fuzzy logic controller. i. Systems, Man and Cybernetics, IEEE Transactions on 20(2), 404–418 (1990).
[13] Sii, H.S., Ruxton, T., Wang, J.: A fuzzy-logic-based approach to qualitative safety modelling for marine systems. Reliability Engineering & System Safety 73(1), 19 – 34 (2001), http://www.sciencedirect.com/science/article/pii/S0951832001000230
[14] Zadeh, L.A.: Fuzzy logic = computing with words. Fuzzy Systems, IEEE Transactions on 4(2), 103–111 (1996).
[15] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of interest point detectors. Int. J. of Computer Vision, 37(2):151-172, 2000.
[16] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Object retrieval with large vocabularies and fast spatial matching. In: Proc. of the IEEE CCVPR, 2007.
[17] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4):600-612, April 2004.